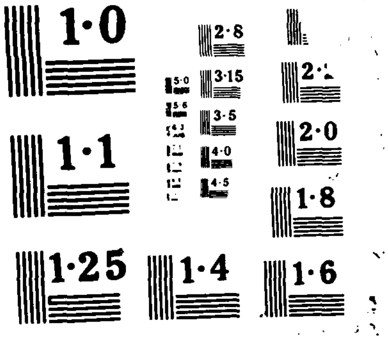
AD-A191 157 PARTIAL LIKELIHOOD ANALYSIS OF TIME SERIES MODELS WITH APPLICATION TO RAI..(U) MARYLAND UNIV COLLEGE PARK DEPT OF MATHEMATICS E V SLUD ET AL. 25 FEB 88 TR-88-11 1/1

UNCLASSIFIED N00014-86-K-0007 F/G 12/1 NL


1·0
2·8
3·15
3·5
4·0
4·5
2·0
1·1
1·8
1·25
1·4
1·6

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(2)

DTIC FILE COPY

AD-A191 157

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER TR88-11 | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) PARTIAL LIKELIHOOD ANALYSIS OF TIME SERIES MODELS, WITH APPLICATION TO RAINFALL-RUNOFF DATA | | 5. TYPE OF REPORT & PERIOD COVERED |
| | | 6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s) Eric V. Slud and Benjamin Kedem | | 8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0007 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS University of Maryland, Mathematics Department, College Park, MD 20742 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, 800 North Quincy Street, Alexandria, VA 22217-5000 | | 12. REPORT DATE February 25, 1988 |
| | | 13. NUMBER OF PAGES 23 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 15. SECURITY CLASS. (of this report) Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

DTIC ELECTE MAR 1 5 1988 S D E

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

logistic regression, time-dependent covariates, partial likelihood, maximum partial likelihood estimation, asymptotic relative efficiency, goodness of fit, hydrological data, level-exceedance

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A general logistic-autoregressive model for binary time series or longitudinal responses is presented, generalizing the discrete-time Cox (1972) model with time-dependent covariates as well as the recent regression models of Kaufmann (1987) for categorical time-series. Since this model is formulated in terms of time-series covariates which are not themselves explicitly modelled, the large-sample theory of parameter-estimation must be justified by means of Partial Likelihood in the sense of Cox (1975),

(continued on reverse side)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

using theoretical results like those of Wong (1986). The large-sample theory also justifies goodness of fit tests analogous to the chi-squared tests of Schoenfeld (1980) and to the tests based on sums of (normalized) squared residuals used in logistic regression. These ideas are illustrated by analysis of a rainfall-runoff hydrological dataset previously analyzed by Yakowitz (1987).

# PARTIAL LIKELIHOOD ANALYSIS OF TIME SERIES MODELS, WITH APPLICATION TO RAINFALL-RUNOFF DATA

by

Eric Slud and Benjamin Kedem
Department of Mathematics
University of Maryland
College Park, MD 20742



MD88-11-ES/BK

TR88-11

March 1988

2/25/88

# Partial Likelihood Analysis of Time Series Models, with Application to Rainfall-Runoff Data

by Eric Slud[1] and Benjamin Kedem[1]
Mathematics Department
University of Maryland
College Park , MD 20742

Abstract: A general *logistic-autoregressive* model for binary time series or longitudinal responses is presented, generalizing the discrete-time Cox (1972) model with time-dependent covariates as well as the recent regression models of Kaufmann (1987) for categorical time-series. Since this model is formulated in terms of time-series covariates which are not themselves explicitly modelled, the large-sample theory of parameter-estimation must be justified by means of Partial Likelihood in the sense of Cox (1975), using theoretical results like those of Wong (1986). The large-sample theory also justifies goodness of fit tests analogous to the chi-squared tests of Schoenfeld (1980) and to the tests based on sums of (normalized) squared residuals used in logistic regression. These ideas are illustrated by analysis of a rainfall-runoff hydrological dataset previously analyzed by Yakowitz (1987).

[1] Research supported by the Office of Naval Research under Contract N00014-86-K-0007 .

# PARTIAL LIKELIHOOD ANALYSIS OF LOGISTIC-AUTOREGRESSIVE AND LONGITUDINAL LOGISTIC MODELS, WITH APPLICATION TO RAINFALL-RUNOFF DATA

by Eric V. Slud[1] and Benjamin Kedem[1]
Mathematics Department, University of Maryland, College Park

**1. Introduction.** A generic description of a great many datasets collected over time is the following: (i) a *univariate* time series $X_t$ [or a set of independent realizations $X^i_t$ of such a series] is of primary interest; (ii) auxiliary vectors $Z_t$ [ or $Z^i_t$ in the case of multiple realizations ] are gathered ***for their value in explaining the behavior of*** $X_t$ in terms of information observable up to time t-1 ; but (iii) no probabilistic models are available concerning the stochastic evolution over time of the explanatory variables $Z_t$ , because the behavior of $\{Z_t\}$ either is too difficult to model effectively or is not of direct scientific interest to justify restrictive assumptions. The scientific problem is then to understand the mechanism by which values $X_t$ , $t=1,\ldots,T$ , arise from the antecedent configuration as of time t-1 . The practical problem in hand may be to ***predict*** , ***forecast*** , or ***classify*** future values $X_t$ from data available up to t-1 ; alternatively, one may wish only to ***estimate*** or ***test hypotheses*** about certain parameters describing the strength of dependence of $X_t$ on antecedent variables $Z_{t-1}$ . In either setting, the problem can be formulated as one of Inference concerning time-invariant parameters $\theta$ of the conditional law of $X_t$ given the observable data $F_{t-1}$ as of time t-1 . Thus it may make sense to perform inferences concerning parameters $\theta$ describing a constant causal structure, even when the "inputs" $Z_{t-1}$

from which $X_t$ arises may be ***nonstationary***. In very different contexts, this idea is familiar both to econometricians (*cf*. the discussion of Wold 1959 concerning ***exogenous variables***) and to biostatisticians (e.g., in the Cox 1972 model with ***time-dependent covariates***). One of the main purposes of the present paper is to argue that the same formulation deserves a more prominent place in time-series analysis.

The class of models which we consider in this paper specializes the previous discussion to the case where the series $X_t$ or $X^i_t$ are ***binary***, i.e., take only the values 1 or 0, and where the conditional laws of $X_t$ or $\{X^i_t\}_i$ given $F_{t-1}$ [the $\sigma$-field generated by all variables $X_s$ and $Z_s$, or $X^i_s$ and $Z^i_s$, for $0 \le s \le t-1$ ] are given by

$$p^i_t(\beta) \equiv P_\beta\{ X^i_t = 1 \mid F_{t-1}, (X^j_t, j \neq i) \} = \{ 1 + e^{-\beta \cdot Z^i_{t-1}} \}^{-1} \qquad (\star)$$

where the real vector parameter $\beta$ has the same dimension d as the covariate vectors $Z^i_t$. Here all vectors are taken to be column-vectors, and $\cdot$ denotes inner-product. Informally, the model $(\star)$ says that the variables $X^i_t$ ( for $i=1,\ldots,n$ and $t=1,\ldots,T$ ) are conditionally independent for different i given $F_{t-1}$, and satisfy a ***logistic regression model*** with respect to the covariates $Z^i_t$. For notational convenience, we take $X_t \equiv X^1_t$ and $Z_t \equiv Z^1_t$ if $n=1$ ; and we take the first coordinate of the covariate vectors $Z^i_t$ to be 1 , so that the intercept-coefficient is $\beta_1$ .

In the model $(\star)$, t is always a discrete time-variable; the

independent time-series replications indexed by $i$ will ordinarily correspond to independent individuals under study; the dichotomous response-variable $X^i_t$ will often indicate that the i'th individual under study failed at time $t$ (on test). In cases where the number $n$ of replicates is $1$, $X_t$ may indicate that $Z_t$ lies inside a certain subset of $R^d$, e.g. that a specified function $V_t \equiv \phi(Z_t)$ is greater than $0$. Finally, the vector $Z^i_t$ of covariates for the i'th replication will be observable by time $t$ but may of course include past values (before $t$) of time-series observables or responses.

The logistic form (*) for the conditional probability that $X^i_t = 1$ is a model assumption which may not always be appropriate. However, there is a simple argument deriving from statistical mechanics which helps to justify logistic regression but does not seem to have been given before by statisticians. The proof of the following assertion is a straightforward exercise on Lagrange multipliers.

*Lemma 1.1.* Suppose that the random variable $X$ takes values $0,1$ and that $p_i \equiv P\{X=i\}$ is restricted in terms of the constants ("covariates") $Y_1 \neq Y_0$ to satisfy the constraint : $p_1 Y_1 + p_0 Y_0 = c$ for a given constant $c$ between $Y_0$ and $Y_1$. The probability $p = p_1 \equiv P(X=1)$ which maximizes the entropy $-p \ln p - (1-p) \ln(1-p)$, is given by $P(X=1) = 1/\{ 1 + exp[-\gamma (Y_1 - Y_0)] \}$, where the constant $\gamma$ is uniquely determined by the constraint.

**2. Examples.** How can time-series models of the type (*) arise ? First, if the covariates $Z^i_t$ for all $t$ and $i$ can be regarded either as design-parameters, or more generally, as known at time $0$, then

*conditionally* given the entire set of covariates, the responses $X^i_t$ follow an ordinary logistic regression model. However, the models (*) are much more interesting and subtle when the covariate-vectors include information about past values $X^i_s$ for $s \le t-1$ or about values $Y^i_s$ of time-series from which $X^i_s$ is derived.

## 2.A. Linear- and Logistic- Autoregressive Models

Consider the case of a single time-series $Y_t$ in terms of which the indicator-series $X_t \equiv X^1_t$ is defined. For example, we may wish to analyze the level-crossing behavior of a continuous-valued time series $Y_t$, in which case we might take $X_t \equiv I_{[Y_t \ge r]}$ for some constant $r$. The model (*) can arise in this context in a very simple way. Suppose that $Y_t$ were an Autoregressive time-series of order $p$, i.e., that

$$\epsilon_t \equiv (Y_t - \gamma_0 - \gamma_1 Y_{t-1} - \cdots - \gamma_p Y_{t-p}) / \lambda$$

is independent of $\{ Y_s : s \le t-1 \}$ with the Logistic density $e^x/(1+e^x)^2$. Then it is easy to check for each fixed $r \in R$ that $X_t \equiv I_{[Y_t \ge r]}$ satisfies (*) with $Z_{t-1} \equiv (1, Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})'$, $d=p+1$, and $\beta \equiv (\gamma_0 - r, \gamma_1, \ldots, \gamma_p)'/\lambda$. In other words, the AR(p) model with logistic errors for $Y_t$ implies (*) for $X_t \equiv I_{[Y_t \ge r]}$ for all $r$. Conversely, if (*) holds for $I_{[Y_t \ge r]}$ for each $r$, with the corresponding coefficients $\beta$ differing only through $\beta_1 = (\gamma_0 - r)/\lambda$, then $Y_t$ is AR(p) with logistic errors. A worthwhile extension of the foregoing is to take

$\epsilon_t \equiv (Y_t - \beta^\star \cdot Z^\star_t)/\lambda$ independent of $\{ Y_s, U_s : s \le t-1 \}$ where $Z^\star_t \equiv (Z_t', U_t')'$, with $Z_t$ defined as above and $U_t$ an auxiliary covariate series observable at the same time as $X_t$. We do not know any models essentially different from these for continuous series $Y_t$, in which $I_{[Y_t \ge r]}$ satisfies (*) for more than one value of r.

The model (*) for $I_{[Y_t \ge r]}$ for a fixed (known) $r$ is a much weaker restriction on $Y_t$ and $Z_t$ than the assumption that $Y_t$ is **AR(p)** with symmetric errors, since it leaves the conditional law of $Y_t - r$ given $\{Z_s : s \le t-1\}$ and $I_{[Y_t \ge r]}$ ***completely unspecified.*** In this way, it is more general than the categorical time-series models of Kaufmann (1987). In Section 5, we describe a dataset where a long continuous-valued series $Y_t$ does seem to obey a model (*) for $I_{[Y_t \ge r]}$ with some fixed values of $r$, but ***does not*** seem to fit an **AR(p)** model.

**2.B. Discrete-time Cox models with time-dependent covariates and multiple event-times.** General models of the type (*) have appeared before in the context of Survival Analysis [ Cox 1975; Andersen and Gill 1982; Arjas and Haara 1987 ]. In this setting, $X^i_t$ is the indicator of the event that the i'th individual under study fails at time t [ or, as in the more usual continuous-time context, has failed before time t ]. The vector $Z^i_t$ of ***time-dependent covariates*** contains all the relevant prognostic information for failure either at time t+1 or in the period (t,t+1]. What is worth emphasizing here is that

models (★) can also accomodate "failures" that may happen more than once to the same individual, such as recurrences of some disease-state like a clinically detectable tumor . (See Gail, Santner, and Brown 1980 for the most notable previous effort to model ***multiple times to tumor***). For such applications, $Z^i_t$ may have the same medical measurement occurring in ***more than one*** component according to how often the individual has previously "failed", i.e., according to the number of values $s=1,\ldots,t$ for which $X^i_s=1$ . The purpose of such a definition is to allow different coefficients $\beta_j$ to operate when previous failures have occurred.

Other applications of models like (★) in the case of relatively few long data-records have been given by Pons and de Turckheim (1985) . For them, $X^i_t$ indicates that the i'th animal (rabbit) under study has eaten something at time $t$ , and $Z^i_t$ includes information about times and amounts of previous feedings. Their method of inference explicitly allows for an additional undetermined (periodic) ***nuisance*** factor $h_0(t)$ not depending on $\beta$ in the expression for the conditional probability $p_t(\beta)$ in (★).

## 3. Partial Likelihood. Regularity conditions for large-sample theory.

The remarkable thing about model (★) is that under mild regularity-conditions on the large-sample behavior of the covariate-processes $(Z^i_t$: $i=1,\ldots,n,\ t=0,\ldots,T)$ as $n\cdot T \longrightarrow \infty$ , there is a numerically stable estimator of $\beta$ which is consistent and asymptotically normal with

easily-estimated variance-covariance matrix. The theoretical apparatus used to justify these statements is the Partial Likelihood of Cox(1975) as developed by Wong (1986) and expounded by Slud(1988). The notion of Partial Likelihood , specialized to our setting, requires precisely what we have so far assumed, namely a model ***for each*** $t$ in terms of the *same* finite-dimensional parameter $\beta$ , for the ***conditional likelihood*** of the observable response data $\{ X^i_t : i=1,\ldots,n \}$ given the ( $\sigma$ field $F_{t-1}$ representing the ) collection of all data observable before time $t$ . The ***Partial Likelihood*** is simply the product, over times $t$ when observations are taken, of these conditional likelihoods. In our setting, the conditional likelihoods are given by (*) together with conditional independence of $X^1_t ,\ldots, X^n_t$ , and the partial likelihood is

$$PL(\beta) \equiv \prod_{t=1}^{T} \prod_{i=1}^{n} [\, p^i_t(\beta)^{X^i_t} \; (1-p^i_t(\beta))^{1-X^i_t} \,]$$

The maximizer $\hat{\beta}^p$ of $PL(\beta)$ is called the Maximum Partial Likelihood Estimator (MPLE) of $\beta$ . Its large-sample behavior under $\beta = \beta_0$ is studied with the aid of the *score-statistic* (d-vector) *process*

$$S_t(\beta) = \sum_{s=1}^{t} \sum_{i=1}^{n} Z^i_{s-1} \, (X^i_s - p^i_s(\beta)) \, , \qquad \beta \in R^d \, , \; t=1,\ldots,T$$

which at $t=T$ is simply the gradient in $\beta$ of *log* $PL(\beta)$ . Since $S_t(\beta)$ is constructed to be a martingale when $\beta=\beta_0$ , the expression

$$I(\beta) = \sum_{t=1}^{T} \sum_{i=1}^{n} Z^i_t \; Z^{i\,\prime}_t \; p^i_t(\beta) \, ( 1-p^i_t(\beta) \, )$$

simultaneously plays the role of Hessian matrix for $-\, log\, PL(\beta)$ with respect to $\beta$ (and is thus an ***information matrix*** for estimating $\beta$), as well as the ***cumulative conditional variance-covariance*** for $S_t(\beta)$ at $\beta=\beta_0$ and $t=T$. The large-sample theory of MPLE's, including the underlying martingale theory, is described in a general setting by Wong(1986) and Slud(1988, Chapter 6), and in successively greater and greater generality in the setting of (*) by Andersen and Gill (1982), Wong (1986), and Arjas and Haara (1987). Related asymptotic results on MPLE's in (*) for large T and fixed n were obtained by Slud (1984), Pons and de Turckheim (1985), Wong (1986), Arjas and Haara (1987), and Slud (1987). Because the ideas underlying Consistency and Asymptotic Normality of MPLE's are very similar in all the papers cited, based on the martingale Central Limit Theorem for $(nT)^{\frac{1}{2}} S_T(\beta_0)$ and stability in probability of $I(\beta)/(nT)$, we do not give any proofs. We content ourselves with some general comments, a precise set of regularity conditions, and statements of results with some small extensions which are helpful in time-series applications.

Since $PL(\beta)$ is almost surely a concave random function of $\beta \epsilon R^d$, with strict concavity under an assumption of asymptotic nonsingularity for $I(\beta)$, there is no need to treat the possibility of multiple MPLE's for large nT. Moreover, concavity removes the need to establish uniform convergence [either a.s. or in-probability] of $I(\beta)/(nT)$ over compact neighborhoods of $\beta$'s when the corresponding pointwise convergence to a continuous concave limit is known (Andersen and Gill 1982, Appendix II). Our regularity conditions, chosen for simplicity rather than utmost generality, are :

**(C.1)** The covariate-vectors $Z^i_t$ almost surely lie in a nonrandom compact subset $\Gamma$ of $R^d$, and the probability measure $P$ governing $\{ X^i_t, Z^i_t : i=1,\ldots,n , t=0,\ldots,T \}$ obeys $(\star)$ with $\boldsymbol{\beta}=\boldsymbol{\beta_0}$ .

**(C.2)** There is a probability measure $\nu$ on $R^d$ for which $\int z\, z' \,\nu(dz)$ is positive-definite, such that under $(\star)$ with $\boldsymbol{\beta}=\boldsymbol{\beta_0}$ ,

$$\frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} I_{[Z^i_{t-1} \in A]} \xrightarrow{P} \nu(A) \quad \text{for all Borel sets } A \subset R^d \text{ as } nT \to \infty$$

Under assumption **(C.2)**, which says that the empirical measure of the set $\{ Z^i_s : 0 \le s < T,\ 1 \le i \le n \}$ converges in a certain sense to a nonrandom measure $\nu$, it follows that for every continuous function $g$ from $R^d$ to $R$ (which is necessarily bounded on the compact support $\Gamma$ of $Z^i_t$ )

$$(nT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{n} g( Z^i_{t-1} ) \xrightarrow{P} \int_{R^d} g(z)\, \nu(dz) \qquad \text{as } n \to \infty$$

From this convergence, it is easy to see that the nonrandom continuous matrix-valued function $\Lambda(\mathbf{b})$ defined for each $\mathbf{b} \in R^d$ as the limit in probability of $I(\mathbf{b})/(nT)$ is given by

$$\Lambda(\mathbf{b}) \equiv \int_{R^d} \frac{e^{\mathbf{b}\cdot z}}{(1 + e^{\mathbf{b}\cdot z})^2} ( z\, z' )\, \nu(dz) \tag{3.1}$$

The matrix $\Lambda(\mathbf{b})$, which is positive semi-definite by inspection for every $\mathbf{b}$ , is also nonsingular at every $\mathbf{b}$ by the hypothesis on $\nu$ . .

**Theorem A.** Under assumptions **(C.1)-(C.2)** , the **MPLE** $\hat{\beta}^p$ is almost surely unique for all sufficiently large nT , is consistent in probability for $\beta_0$ as $nT \to \infty$, and satisfies $(nT)^{\frac{1}{2}} (\hat{\beta}^p - \beta_0) \xrightarrow{D} N(\underline{0}, \Lambda(\beta_0)^{-1})$ . In addition, $(nT)^{\frac{1}{2}} (\hat{\beta}^p - \beta_0) - (nT)^{-\frac{1}{2}} \Lambda(\beta_0)^{-1} S_T(\beta_0) \xrightarrow{P} 0$ as $nT \to \infty$ .

The large-sample behavior of $\hat{\beta}^p$ described here is based on the assumed stability of information $I(\beta)/(nT)$ ***per observation*** . The only real novelty in our presentation is that a single theoretical result is made to encompass both the cases where n becomes large and those where T does. Other possible limiting distributional behavior is possible in so-called ***nonergodic*** cases (Basawa and Scott 1983) where normalized information converges in probability to a random limit . The reason for referring to **(C.2)** as an ergodic setting is that it is follows from the Birkhoff Ergodic Theorem either in the case of independent identically distributed processes $(Z^i_t: 0 \le t \le T)$ for $i=1,\ldots,n$ with T fixed , or of an ergodic and stationary process $(Z^1_t,\ldots,Z^n_t)$ in $t=0,1,\ldots$ with n fixed.

In attempts to fit (*) to data, it is crucially important to be able to assess the quality of fit of the model with parameters $\beta$ in terms of the ***residuals*** $X^i_t - p^i_t(\beta)$ . These are precisely the residuals which one uses in ordinary logistic regression, and two sorts of goodness-of-fit statistics based on them are of particular value. The idea of the first is to check the observed number of responses $X^i_t = 1$ corresponding to $Z^i_{t-1}$ within each of a number of covariate-defined cells against the

cumulative one-step-ahead predicted number. The resulting chi-squared statistic is closely related to the one given in a Survival context by Schoenfeld (1980). Our second type of statistic is a sum of squared normalized residuals. The theory underlying both statistics is contained in the following two theorems, which follow readily from Theorem A and several applications of the Martingale (multivariate) Central Limit Theorem (as given for example in Andersen and Gill 1982, Appendix I). In both Theorems, it is crucial that the set of possible $p^i_t(\beta)$ values for all $i$ and $t$ and all $\beta$ in a neighborhood of $\beta_0$ lie in a nonrandom compact subset of $(0,1)$. This fact follows from **(C.1)**.

<u>**Theorem B.**</u> Let $C_1, \ldots, C_k$ be a partition of $R^d$ into measurable sets, and define $N_j$ and $E_j(\beta)$ for $j=1,..,k$ through the formulas

$$N_j \equiv \sum_{i=1}^{n} \sum_{t=1}^{T} I_{[Z^i_{t-1} \in C_j]} X^i_t , \quad E_j(\beta) \equiv \sum_{i=1}^{n} \sum_{t=1}^{T} I_{[Z^i_{t-1} \in C_j]} p^i_t(\beta)$$

Put $\underline{N} \equiv (N_1,\ldots N_k)'$ and $\underline{E}(\beta) \equiv (E_1(\beta),\ldots,E_k(\beta))'$. Then for fixed $k$,

$$(nT)^{-\frac{1}{2}} \left( (\underline{N} - \underline{E}(\beta_0))' , (\hat{\beta}^P - \beta_0)' \right)' \xrightarrow{\mathcal{D}} \mathcal{N}(\underline{0}, \Sigma)$$

where $\Sigma$ is a square $d+k$ dimensional matrix of the form $\begin{pmatrix} A & B' \\ B & D \end{pmatrix}$. Here $A$ is $k\times k$ and diagonal with j'th diagonal element

$$\sigma_j^2 = \int_{C_j} \frac{e^{\beta_0 \cdot z}}{(1+e^{\beta_0 \cdot z})^2} \, \nu(dz) , \quad D = \Lambda(\beta_0)^{-1} \text{ is } d\times d , \text{ and the j'th}$$

column of $B$ is given by $\displaystyle\int_{C_j} \frac{e^{\beta_0 \cdot z}}{(1+e^{\beta_0 \cdot z})^2} \, Dz \, \nu(dz)$. In addition,

$$(nT)^{-\frac{1}{2}} \left( \underline{E}(\hat{\beta}^P) - \underline{E}(\beta_0) \right) - (nT)^{-\frac{1}{2}} B' D^{-1} (\hat{\beta}^P - \beta_0) \xrightarrow{P} \underline{0} \quad \text{as } nT \to \infty .$$

It follows that as $nT \to \infty$, both the statistics $\sum_{j=1}^{k} (N_j - E_j(\beta_0))^2 / \sigma_j^2$ and $(\underline{N} - \underline{E}(\hat{\beta}^P))' (A - B' D^{-1} B)^{-1} (\underline{N} - \underline{E}(\hat{\beta}^P))$ are asymptotically $\chi_k^2$-distributed.

<u>Theorem C.</u> Under hypotheses **(C.1)** and **(C.2)** , let $a \geq 0$ be fixed arbitrarily, and define $v(i,t) \equiv p^i_t(\beta_0)\ (1 - p^i_t(\beta_0))$ for all $i, t$ . Then

$$\frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \left[ \frac{(X^i_t - p^i_t(\beta_0))^2}{v(i,t)^a} - [v(i,t)]^{1-a} \right]}{\left\{ \sum_{i=1}^{n} \sum_{t=1}^{T} \{v(i,t)\}^{1-2a} \cdot \{ 1 - 4\, v(i,t) \} \right\}^{\frac{1}{2}}} \xrightarrow{D} N(0,1) \quad \text{as } nT \to \infty.$$

The result of Theorem C is equally valid when the residuals and $v(i,t)$ are calculated with $\beta_0$ replaced by $\hat{\beta}^P$. However, in applications where predicted response-probabilities $p^i_t(\beta_0)$ can get very close to 0 or 1 , the behavior of the normal deviates in Theorem C may not be very reliable .

## <u>4. Asymptotic-efficiency calculations. Extensions to other link-functions.</u>

How efficient is a data analysis based on (*) and the Partial Likelihood when a fully specified model for $(X^i_t, Z^i_t)$ is available ? While no general answer to this question is possible, we can calculate the

asymptotic relative efficiency in the **AR(p)** Example of § 2.A , where for a fixed $r \geq 0$ ,

$$X_t \equiv X^1_t \equiv I_{[Y_t \geq r]}, \qquad Z_{t-1} \equiv Z^1_{t-1} \equiv (1, Y_{t-1}, \ldots, Y_{t-p})'$$

$$\beta \equiv ((\gamma_0 - r)/\lambda,\ \gamma_1/\lambda,\ \ldots,\ \gamma_1/\lambda)', \qquad \epsilon_t \equiv (\, Y_t - \beta \cdot Z_{t-1}\,) / \lambda$$

and $\epsilon_t$ is assumed to be independent of $F_{t-1} \equiv \sigma(\, Z_s : s<t\,)$ and to have the logistic density $f(x) \equiv e^x/(1+e^x)^2$ . Taking the initial data $Z_0$ to be given, one can easily derive the likelihood $L(\beta)$ for the observed data $\{\, Y_t \,\}_{t=1}^{T}$ as well as the limiting information about the parameter vector $\beta$ per observation. This information-matrix, equal to the inverse of the asymptotic variance matrix for the maximum-likelihood estimator of $\beta$ when the true parameter value is $\beta_0$ , is given by

$$I(\beta_0) \quad = \quad E_{\beta_0}\{\, \frac{e^{\epsilon}}{(\,1+e^{\epsilon})^2}\, Z\, Z' \,\} \quad = \tfrac{1}{3}\, E_{\beta_0}\{\, Z\, Z' \,\}$$

where $E_{\beta_0}$ denotes expectation when the random (p+1)-vector $Z$ has the ***stationary*** distribution for $Z_t$ (assumed to exist) as $t \to \infty$ , and $\epsilon$ is a logistic random variable independent of $Z$ . The corresponding "partial likelihood" information-matrix arising in Theorem A above was derived in (3.1) as

$$I^p(\beta_0) \;\equiv\; \Lambda(\beta_0) \;=\; E_{\beta_0}\{\, \frac{e^{\beta_0 \cdot Z}}{(1+e^{\beta_0 \cdot Z})^2}\; Z\, Z' \,\} \;=\; E_{\beta_0}\{\, f(\beta_0 \cdot Z)\, Z\, Z' \,\}$$

Since the function $f(x)$ has maximum value $\frac{1}{4}$ , it follows immediately from the last two formulas that for any vector $b \in R^{p+1}$ ,

$$b' \, I^p(\beta_0) \, b \;\le\; \tfrac{3}{4} \; b' \, I(\beta_0) \, b$$

Thus, any scalar parameter derived from $\beta$ can be estimated with asymptotic relative efficiency (ARE) ***at best*** $\frac{3}{4}$ via the Partial-Likelihood Logistic-Regression method as compared with a maximum-likelihood **AR(p)** analysis. Of course, the **ARE** could (for some **AR(p)** models) be much worse than $\frac{3}{4}$, but it should be somewhat reassuring that the cases where relative efficiency is worst are the cases where the predictors $p_t(\beta_0)$ spend most of their time close to 1 or 0, i.e. the cases where prediction is very good !

The foregoing analysis of model (*) has many possible generalizations. One that is especially worth pursuing is to allow the predictors $p^i_t(\beta)$ in model (*) to take some other parametrically specified form $F(\beta \cdot Z^i_{t-1})$ as a function of a linear expression in the explanatory variables $Z^i_{t-1}$. One choice for the "link" function $F$, which leads to methods bearing the same relation to ***probit regression*** as our previous analysis does to logistic regression, is the standard normal distribution function $F \equiv \Phi$. The model from which conditional likelihoods are constructed is now

$$P_\beta\{ \; X^i_t = 1 \mid \mathcal{F}_{t-1} \, , \, (X^j_t \, , \, j \neq i) \; \} \;=\; \tilde{p}^{\,i}_t(\beta) \;\equiv\; \Phi(\beta \cdot Z^i_{t-1}) \qquad (**)$$

Virtually every aspect of our analysis of (*) has an analogue for (**). The **AR(p)** time-series examples of (**), with $X_t \equiv I_{[Y_t \ge r]}$, will now have Gaussian error-distribution. The proofs of Theorems analogous to A, B, and C are very similar to those in the logistic case, except for some extra complication since the log Partial Likelihood will no longer

be a.s. concave for finite data-samples. Finally, the **ARE** calculations of this Section can be carried out for the **AR(p)** examples of (**) and lead to formulas

$$I(\beta_0) = E_{\beta_0}\{ Z Z' \} \quad , \qquad I^P(\beta_0) = E_{\beta_0}\{ Z Z' \frac{\phi^2(\beta_0 \cdot Z)}{\Phi(\beta_0 \cdot Z)(1-\Phi(\beta_0 \cdot Z))} \}$$

where $Z$ now has the stationary distribution of $Z_t$ in an **AR(p)** model with ***normal*** errors, and $\phi$ is the standard normal density. The highest possible value for **ARE** in estimating scalar parameters related to $\beta$ is now $2/\pi$ , and again the **ARE** can be much worse than that in cases where prediction is very accurate.

5. **Application to Rainfall-runoff data.** We now describe the results of applying model (*) to the analysis of a hydrological dataset which has previously been analyzed by quite different methods. The data, which is described by Yakowitz (1987), consist of daily measurements by the National Weather Service of rainfall and runoff in the Bird Creek Ohio watershed. Data were collected for a 13-15 week period during each of the years 1939-1964. The purpose of collecting such data seems partly to have been to develop models for the prediction of flooding. Thus we regard daily runoff $Y_t$ as the (continuous-valued) response variable, with past runoff together with current and past values of rainfall $R_t$ playing the role of explanatory variables. Since flooding is of interest, it is natural to try to understand the relationships between level-exceedances $X_t = I_{[Y_t \geq r]}$ and the explanatory variables. Our use of this example is not intended primarily to develop a formula for prediction, but rather to illustrate how models (*) may be estimated

and shown to pass tests of adequacy in a particular situation where the linear autoregressive model turns out not to be appropriate.

For our data analysis, we split the 26 years of data into a training set [ the years 1939-48, consisting of 1031 rainfall-runoff pairs ] and a testing set [ the years 1949-64, consisting of 1691 observation-pairs ]. Since the models we contemplated involved explanatory variables defined from $(Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, R_t, R_{t-1}, R_{t-2}, R_{t-3})$, our covariate-data for the first four response-variables $X_t$ in each year were incomplete. Deleting these observations led to a complete dataset of 991 observations for model-fitting and 1627 observations for testing model adequacy. The cutoff threshold $r$ used in defining the indicator response $X_t$ was chosen to be 1 or 3 (cubic ft/sec.) These levels respectively corresponded to 244 and 56 positive responses [i.e., values $Y_t$ above $r = 1, 3$ ] out of 991 in the fitting-dataset, and to 401 and 87 out of 1627 in the test data.

After some preliminary fitting [via maximum partial-liklihood], plotting of residuals, and computation of partial likelihood-ratio statistics for the 1939-48 data, we found that the important covariates $Z_t$ in model (*) for $I_{[Y_t \geq r]}$ (r=1 or 3) were $R_t$, $Y_{t-1}$, $R_t{*}Y_{t-1}$, $R_{t-1}$, $R_t{*}R_{t-1}$, $R_{t-2}$, $R_{t-1}{*}R_{t-2}$, $Y_{t-2}$, and $Y_{t-1}{*}Y_{t-2}$. Although the mechanics of soil-saturation naturally suggest that interaction-terms involving successive days' rainfall and runoff should play some role in explaining future runoff, previous linear — as opposed to logistic — regression analyses do not seem to have included such terms. As it turned out, not all the terms listed above had significantly large coefficients for each threshold $r$, but no *other* terms appeared to play

of a role. In particular, our attempts to discover covariates to account for year-to-year differences in runoff — and such differences do appear to be real — did not produce new covariates worth including in the model. For interpretability of coefficients, we continued to include separate terms corresponding to each significant interaction-term.

The values $\hat{\beta}_i^P$ and standardized values $\hat{\beta}_i^P/[\mathrm{Var}(\hat{\beta}_i^P)]^{\frac{1}{2}}$ of the fitted coefficients $\beta$ in model (★) are exhibited in Table 1, for each of the cases $X_t \equiv I_{[Y_t \geq r]}$ with $r=1$ and $r=3$. In each case, the 10-dimensional covariate vector is $Z_t' \equiv$

$$(1, R_t, Y_{t-1}, R_t{\star}Y_{t-1}, R_{t-1}, R_t{\star}R_{t-1}, R_{t-2}, R_{t-1}{\star}R_{t-2}, Y_{t-2}, Y_{t-1}{\star}Y_{t-2}) \tag{5.1}$$

The most important covariates are : the intercept term , the interaction terms $R_t{\star}Y_{t-1}$ , $R_{t-1}{\star}R_{t-2}$ , $Y_{t-1}{\star}Y_{t-2}$ , as well as the obviously important covariates $R_t$ , $Y_{t-1}$ , and $R_{t-1}$ and/or $R_{t-2}$ . The log-likelihoods for the fitted models (on the 991 training-observations) were -142.7 for $r=1$ and -71.5 for $r=3$ .

Motivated by the results of Yakowitz (1987), we tried also to fit models in which the linear residuals $Y_s - \hat{\beta}{\cdot}Z_s \equiv \epsilon_s$ for $s=t-1$ and $s=t-2$ from a preliminary fit would serve as additional covariates [i.e. as additional components of an augmented Z-vector. ] From the perspective of partial likelihood-ratio tests ***performed on the training data alone***, the coefficients of these additional covariates were on the borderline of significance ; the maximized partial likelihoods [ for 12-dimensional covariate-vector vs. the previous 10-dimensional covariate-

vector (5.1) ] were respectively -140.2 for the case $r=1$ and -68.2 for $r=3$ . [Thus twice the log partial likelihood ratios for the two cases are 5.0 and 6.6 , which should be compared with the .05 percentage-point of 5.99 for the $\chi_2^2$ -distribution . The validity of partial-likelihood ratio tests in this context follows from Theorem A by standard arguments with the asymptotic multivariate-normal distribution of $\hat{\beta}^p$ .] The slight advantage of the extra two covariates disappeared when we applied the ten- and twelve-covariate models to the testing-dataset (the 1627 observations for years 1949-1964). Therefore we confine our further discussion to the fitted models (*) based on covariates (5.1) alone .

The previous attempts surveyed by Yakowitz (1987) to fit models to rainfall-runoff data involved ***linear*** (auto-) regressions, with or without moving-average error terms. As we have seen above in Section 2 , if the error-terms were approximately logistically distributed then we should expect the logistic autoregressive models fit with different thresholds $r$ to share the same important covariates and coefficients (other than the intercept-term). Even a cursory inspection of the fitted coefficients in Table 1 shows that this is not the case for our rainfall-runoff data. This is indirect evidence for inadequacy of a linear-regression model. Indeed, after numerous linear-regression fits with different sets of covariates, we convinced ourselves that prediction of $I_{[Y_t \geq r]}$ was much less good with linear than with logistic models.

To begin to assess model adequacy for the logistic models (*) fitted to the 1939-1948 rainfall-runoff data, we calculated goodness-of-fit statistics as described in Theorems B and C of Section 3 , replacing

the (unknown) parameters $\beta_0$ in those statistics by their estimators $\hat{\beta}^p$ based on the 1939-1948 data . The k (=27) cells used to define the $\chi^2_k$ statistic of Theorem B were based on arbitrary cutoffs 0.004 and 0.008 for $R_t$ , 0.01 and 0.02 for $R_t+R_{t-1}$, and 0.5 and 1.0 for $Y_{t-1}$ , and the variances $\sigma^2_j$ of Theorem B were estimated from the data as described in Table 2 below. The goodness-of-fit statistics were calculated (using the same estimated $\hat{\beta}^p$ ) both on the training dataset (1939-1948) and on the testing-dataset (1949-1964), with illuminating results. The goodness-of-fit chi-squared statistics $\chi^2_B$ based on a

---

Table 1. Coefficients and Standardized Values for Model (*) fitted to Rainfall-Runoff Data for Years 1939-48 with covariates (5.1) .

| | r=1 case | | r=3 case | |
|---|---|---|---|---|
| Coefficient index | $\hat{\beta}_i^p$ | standardized | $\hat{\beta}_i^p$ | standardized |
| 1 | -5.933 | -8.10 | -5.655 | -9.59 |
| 2 | 72.49 | 1.81 | 92.86 | 4.48 |
| 3 | 3.633 | 5.63 | -0.397 | -1.97 |
| 4 | 302.2 | 3.97 | 99.37 | 4.39 |
| 5 | 14.68 | 0.44 | 121.46 | 4.20 |
| 6 | 6985. | 1.59 | -3139 | -2.14 |
| 7 | -59.87 | -2.22 | 43.97 | 0.94 |
| 8 | -11504 | -2.56 | -2422 | -1.88 |
| 9 | -0.630 | -0.87 | -0.42 | -0.82 |
| 10 | 1.022 | 2.04 | 0.128 | 1.87 |

decomposition of the covariate-space into $3^3 = 27$ cells should be compared with a quadratic-form in multivariate-normal random variables which is stochastically smaller than $\chi^2_{27}$ on the training-data [due to having estimated $\beta_0$ from the same data] and stochastically larger than $\chi^2_{27}$ on the testing-data [since the estimated $\hat{\beta}^p$ is approximately independent of the deviations $\underline{N} - \underline{E}(\beta_0)$ for the testing-data] . In the notation of Theorem B , the expectation of $\chi^2_B$ on the training-data is $k - \Delta$ and its expectation on the test-data is $k + \Delta$ , where $\Delta = \sum_{j=1}^{k} (B'D^{-1}B)_{jj} / \sigma^2_j$ . In our example, with $k=27$, the (empirically estimated) value of $\Delta$ turns to be ~~0.4~~ 5.8 when $r=1$ and ~~7.5~~ 3.0 when $r=3$ . The different behavior of $\Delta$ for the different thresholds $r$ appears to be due to the very different predicted response-probabilities $p_t(\hat{\beta}^p)$ which arise in the models for different $r$ with a given covariate-covariate-vector $Z_t$ . Table 2 indicates that according to all three of our goodness-of-fit statistics, the models (*) with $X_t = I_{[Y_t \geq r]}$ for both $r=1$ and $r=3$ are adequate for the training data ***and*** that the corresponding models with coefficients fitted to the training data (years 1939-48) are seriously inadequate for the the test-data (years 1949-64). This need not be too distressing, since there do seem to be noticeably different rainfall-runoff patterns, including differences in overall runoff levels, from year to year, although the differences are not systematic enough to be easily encoded into additional covariates . Some care is needed in interpreting the results using the statistics

$\sum_{t=1}^{T} (X_t - p_t(\hat{\beta}^P))^2 / [\, p_t(\hat{\beta}^P)(1-p_t(\hat{\beta}^P))\,]$ from Theorem C , since many of the logistic prediction-probabilities $p_t(\hat{\beta}^P)$ based on either dataset are quite close to zero or one. For this reason , the results in Table 2 for $W_{C,0}$ are probably more reliable than those for $W_{C,1}$ .

**Table 2 . Goodness-of-fit statistics on training- and testing- data for Model (*) with coefficients displayed in Table 1.** The chi-squared statistic $\chi_B^2$ is defined as in Theorem B with $\boldsymbol{\beta_0}$ replaced by $\hat{\beta}^P$ , with $\sigma_j^2$ estimated by $\hat{\sigma}_j^2 \equiv T^{-1} \sum_{t=1}^{T} I_{[Z_{t-1} \in C_j]}\, p_t(\hat{\beta}^P)\,(1-p_t(\hat{\beta}^P))$ , and with 27 partition-cells $C_j$ defined as the intersections of all sets $R_t \in [0,.004]$ , $(.004,.008]$ , or $(.008,\infty)$ ; $R_{t-1}+R_t \in [0,.01]$ , $(.01,.02]$, or $(.02,\infty)$ ; and $Y_{t-1} \in (0,.5]$ , $(.5,1]$ , or $(1,\infty)$ . The statistics $W_{C,a}$ for $a = 0$ and $a=1$ are the asymptotically normal deviates defined in Theorem C , with $\boldsymbol{\beta_0}$ replaced by $\hat{\beta}^P$. In all statistics, estimators $\hat{\beta}^P$ and $\hat{\sigma}_j^2$ are calculated from ***training-dataset*** (**'39-'48**).

| Statistic | Training-data (991 obs.) | | Test-data (1627 obs.) | |
|---|---|---|---|---|
| | r=1 | r=3 | r=1 | r=3 |
| $\chi_B^2$ | 31.0 | 7.47 | 98.8 | 52.6 |
| $W_{C,0}$ | 0.47 | -0.03 | 7.4 | 2.4 |
| $W_{C,1}$ | -0.01 | 1.14 | 44.9 | 18.8 |

A further interesting conclusion from fitting models (*) to the 1939-1948 rainfall-runoff data of Yakowitz (1987) is that while these models appear to fit reasonably well both with the level-crossing thresholds $r=1$ and $r=3$, they do not in this example appear compatible with a single linear model for runoff. In this sense, we have a real example of (*) different from the examples of Section 2.A which derive from AR(p) models with logistic errors.

## 6. References


Andersen, P. and Gill, R. (1982) Cox's regression model for counting processes: a large sample study. *Ann. Statist.* 10, 1100-1120.

Arjas, E. and Haara, P. (1987) A logistic regression model for hazard: asymptotic results. *Scand. Jour. Statist.* 14, 1-18.

Basawa, I. and Scott, D. (1983) Asymptotically Optimal Inference for Non-Ergodic Models. Lecture Notes in Statist. 17. New York: Springer-Verlag.

Cox, D.R. (1972) Regression models and life tables. *Jour. Roy. Statist. Soc.* B 34, 187-202.

——— (1975) Partial likelihood. *Biometrika* 62, 69-76.

Gail, M., Santner, T., and Brown, C. (1980) An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* 36, 255-266.

Kaufmann, H. (1987) Regression models for nonstationary categorical time series: Asymptotic estimation theory. *Ann. Statist.* 15, 79-98.

Pons, O. and De Turckheim, (1985) Cox's periodic regression model. Preprint.

Schoenfeld, D. (1980) Chi-squared goodness-of-fit for the proportional hazards regression model. *Biometrika* 67, 145-153.

Slud, E.V. (1984) Multivariate dependent renewal processes. *Adv. Appl. Prob.* 16, 347-362.

——— (1987) Efficiencies of partial-likelihood-based inferences concerning survival regression models. Submitted for publication.

——— (1988) Martingale Methods in Statistics. Forthcoming book.

Wold, H. (1959) Ends and means in econometric modelling, pp.355-434 In: Probability and Statistics: The Harald Cramer Volume, ed. U. Grenander. New York: John Wiley.

Wong, W.H. (1986) Theory of partial likelihood. *Ann. Statist.* 14, 88-123.

Yakowitz, S. (1987) Nearest-neighbour methods for time series analysis. *Jour. Time Ser. Anal.* 8, 235-247.

END
DATE
FILMED
5-88
DTIC